Public Opinions on Russia-Ukraine War in 2022

Sentiment Analysis

Image：Adobe

Prepared by: Group 16

December 2nd, 2022

Table of Content

Introduction

The controversial conflicts between Russia and Ukraine have initiated intense discussions on social media, with some people holding positive attitudes toward the war and others being depressed or remaining neutral. The objective of this report is to use sentiment analysis to algorithmically analyze people's attitudes towards the Russia-Ukraine War and the potential reasons behind their thoughts based on their posted tweets and then suggest effective ways to positively improve the international image of Ukraine. Four machine learning models were applied to estimate the attitudes of the tweets, and the best one can result in over 95% accuracy.

Summary

Suggestion to improve Ukraine international image:

1. Advocating peace is the trend in nowadays society, and the public understands how citizens will suffer in wars. Showing how people are living under this circumstance can enable others to understand that the situation is serious and non-negligible. Ukraine may receive more support, help or attention from other countries or individuals.

2. Apart from the support of individuals, calling on international parties or organizations who take peace or equity as the core is considerable. The organizational interventions may provide Ukraine great assistance by strengthening communication or helping to reach a consensus.

3. Propagandize the strong relationships with the supporters or ally countries. As the public also pays attention to the political stand of other countries in this war. Showing affiliation with the countries who are holding great international influence can also improve the presence of Ukraine and inspire the citizens at the same time.

4. Publicizing the negative effect of the war from an international perspective. Considering the economic globalization, the conflict can result in economic inflation world widely.

5. The socially vulnerable groups, refugees, women, or children will suffer more in this war. Some countries surrounding Ukraine have provided safe shelters and humanitarian assistance to millions of refugees, expressing appreciation and thanks may also improve the international image of Ukraine for the supporters.

1 | Sentiment Modeling

1.1 Data preprocessing

The texts from the internet are complicated and massive. Before implementing appropriate models, transforming the text data into simple and clean words which are readable for computers is essential. Irrelevant or less-informative content will be removed as well to save time in applying the algorithms.

Mentions that can be frequently found in tweets are removed since it does not relate to the sentiments of the posters. Hashtags are separated from the context of the tweets as they are more likely to represent the topics of the tweets rather than attitudes. Besides, punctuations, emojis, and words that are meaningless i.e., ‘are’, ‘he’ and numbers are removed. After removing less informative parts, the long text of tweets is split into words and each word is treated as a unit which will be used as a single feature in the further models. Words are also being transformed into their base words which allow us to group together words with similar meanings into one word. After the procedures mentioned above, the text data is condensed and only the significant parts are left. Now, in order to implement the models, features need to be extracted from the text in a numerical way to be better understood and estimate the sentiments of the tweets’ authors by models.

TF-IDF is a way of extracting features from text for use in modeling, which measures the importance of each word. Since there are over 10000 unique words in the document, the most important 2000 words are selected. The result gives the TF-IDF score corresponding to the top 2000 words for each tweet post. The higher the TF-IDF score the more important the word is.
After extracting the numerical information of the importance of each word to the dataset, we can teach algorithms to identify positives and negative and make estimations.

1.2 Model implementations

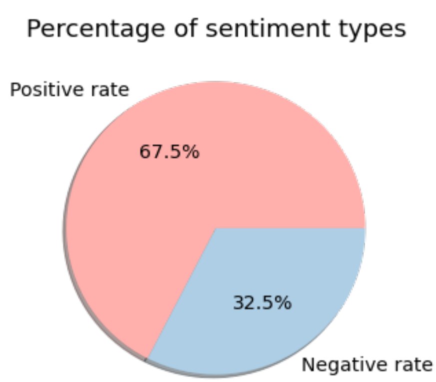


Figure 1-1

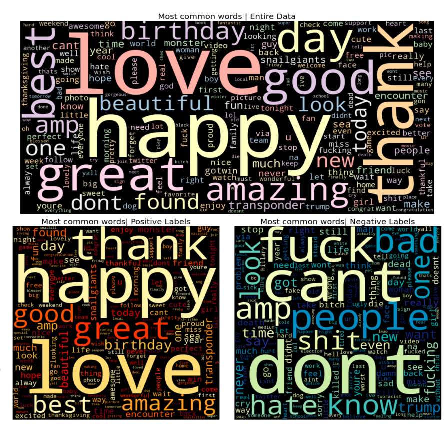
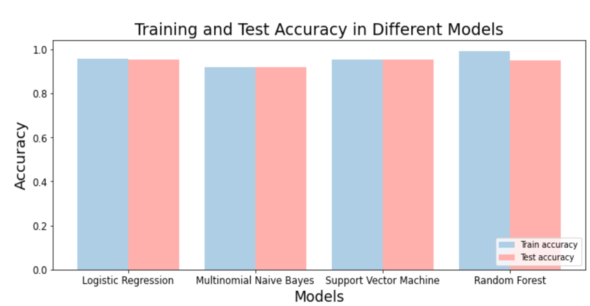


Figure 1-2

A dataset containing more than 550,000 counts of tweets collected directly from the web is used to train and validate the sentiment models. These tweets along with the polarity labels indicating positive or negative are randomly chosen, and not all of them are related with the opinion on the war. Among the random tweets, an approximate of 67.5 % are classified as positive, with the most frequent words are ‘happy’, ‘birthday’, ‘love’, and the rest 32.5 % are attributed to negative where ‘look’, ‘fuck’, ‘heat’ appeared much often, as shown in Figure 1-1 and Figure 1-2.
Next, four classification algorithms, logistic regression, Naive Bayes, SVM, and Random Forest, are trained and compared in the bar plot (figure 1-3) in order based on the 70% whole dataset, and the rest are used to test their performance respectively. As the samples with different labels are not much imbalanced, accuracy is considered as a metric of performance on category tweets.



The best model with the highest accuracy is Logistic regression with C=20, solver =' saga' and it has **95.43 percent** accuracy.

Figure 1-3

Logistic regression

Logistic regression estimates the probabilities of an event occurring, in our case positive sentiment or negative sentiment. The model is trained at first on a dataset with the labeled sentiment, so the model will know what types of features normally occur in positive or negative tweets and the model will learn how to estimate tweets with unknown labels. Based on the number of occurrences of each word, logistic regression models will tell the probabilities of each event occurring and whichever gets the higher probability, the tweets will be assigned to that class. However, since 2000 features were chosen, with that many features, overfitting might occur, and the models need to be penalized to prevent the problem. By penalizing the logistic model, the model will automatically assign importance to each feature, and it will ignore the features with 0 importance. The harder it is to penalize the model; the more features will be dropped. By tuning the inverse of the penalization terms C, the C which leads to the highest accuracy will be chosen. Besides penalization, logistic models are also about optimization, minimizing the error. The hyperparameter solver is the algorithm that is applied by the models to compute the optimized point. Some solvers work better for large datasets, and some are good for small datasets. By trying different solvers, the best optimizing algorithm will be chosen based on the training dataset for the models.

Naive Bayes

Naive Bayes estimates probabilities of whether the authors of the tweets have a positive or negative sentiment on the condition of the importance of the words. One thing that makes Naive Bayes naive is that the model assumes that each feature is independent of each other. For instance, a fruit may be considered to be a watermelon if it is green, round and about 30 cm in diameters. Naive bayes consider all the features that contribute independently of that fruit to be watermelon and ignore the correlation between the colors, roundness, and the diameters. But Naive Bayes is time saving especially on large datasets and the accuracy of the model is quite high.

Linear Support Vector Machine

Linear Support Vector Machine (Linear SVC) is an algorithm that attempts to find a separator that separates negative sentiment samples and positive sentiment samples. Same as logistic regression, Linear SVC needs to be penalized to improve the model accuracy. Hyperparameter C is chosen for tuning. As the result, the model with C=1 perform the best, with test accuracy of **95.38%**

Random Forest:

Random Forest is an algorithm that combines multiple decision trees. While a decision tree attempts to build a flowchart leads to an outcome. As the number of trees grows, the running time of the algorithm increases. The hyperparameter chosen to tune is the number of trees. As a result, the model with 100 trees performs the best, with test accuracy of **95.17%**.

2 | Sentiment Classification

The two datasets chosen here to contain scraped tweets related to Russia's war in Ukraine, which will be used on the four models built from the last part and discussed respectively. As one of the pre-trained sentiment classification models in

Python, SentimentIntensityAnalyzer with the Vader function can classify the tweets as positive, neutral, or negative. The labels resulted by SentimentIntensityAnalyzer are considered as criterions to make evaluation of the models.

2.1 Sentiment classification, Dataset 1

The first pick of the datasets contains 23179 samples, after data preprocessing and model implementation, the pre-trained model results in 8647 negative, 8838 positive and 5694 neutral. As each model built above returns outputs in negative or positive only, for a better comparison with the 3-class criteria, we decide to label the outputs according to their probability to positive. The boundaries we have tried are indicated in the Table 2-1 below.

Table 2-1:

Probability to positive	0 <= p < 0.4	0.4 <= p <= 0.6	0.6 < p <= 1
Label	Negative	Neutral	Positive

As a result, the performance of neither of the four models is desirable, as each has an accuracy just around 53%. The possible reason behind the situation can be attributed to the ambiguous definition of ‘neutral’ across different individuals. Additionally, compared with the neutral opinions, the insight will be gained more through the tweets with negative or positive sentiment, so we decide to evaluate the performance of the two polarity opinions only, and the ratios are shown in the pie chart (Figure 2-1). The result here displays an improvement on accuracy by more than 20% for each of the four models, indicating the models built can do a decent job on predicting positive or negative sentiment. The accuracies are shown in Table 2-2. Word clouds for both, Positive, Negative samples, and the ratios are displayed in Figures 2-2 as well.

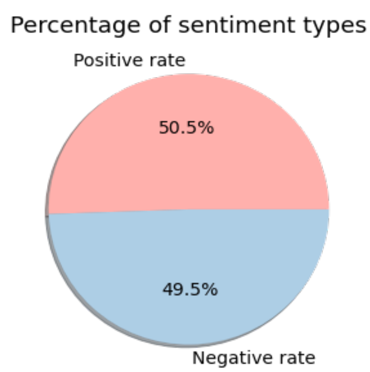


Figure 2-1

Figure 2-2

2.2 Sentiment classification, Dataset 2

To improve the reliability of the precision, another 20,000 tweets were scraped using snsrarpe which are related with the war and posted just between June 1 to November 15, 2022. Drop the samples predicted as neutrals by SentimentIntensityAnalyzer, the remaining are classified to positive or navigate, and the ratios are shown in the pie chart (Figure 2-3). The accuracy of the prediction for the two classes are shown in the Table 2-2 below, and the word clouds are in Figure 2-4.

Table 2-2: Accuracy

Data Set	Labels	Logistic Regression	Naive Bayes	SVM	Random Forest

0819_UkraineCombinedTweetsDeduped.csv	Negative/**Neutral**/Positive	**54.72%**	**53.75%**	NA	**52.47%**
0819_UkraineCombinedTweetsDeduped.csv	Negative/ Positive	**72.65%**	**72.33%**	**72.66%**	**69.35%**
war_tweets.csv	Negative/ Positive	**84.78%**	**79.63%**	**84.45%**	**79.65%**

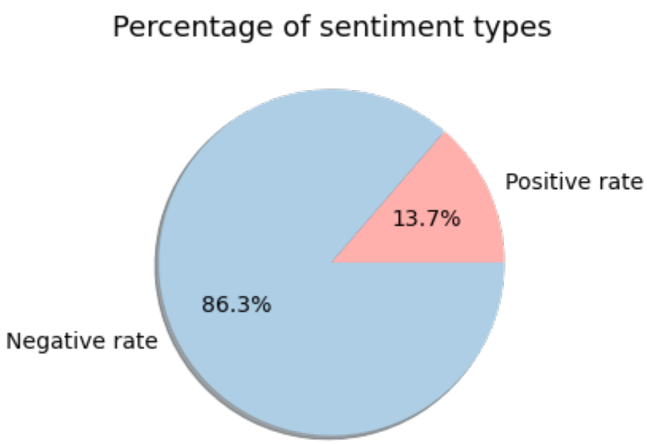


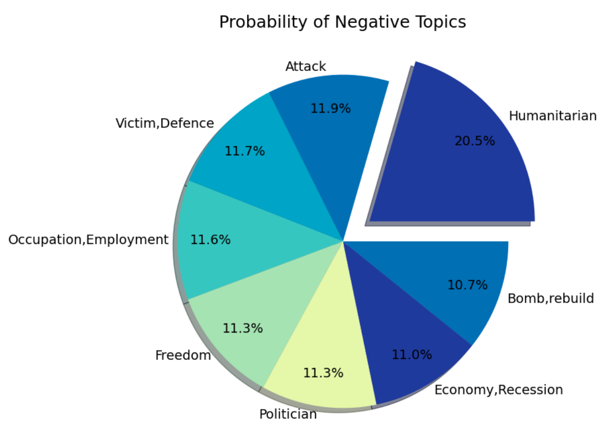
Figure 2-3

Figure 2-4

3 | Topics Identification via Machine Learning, Storytelling

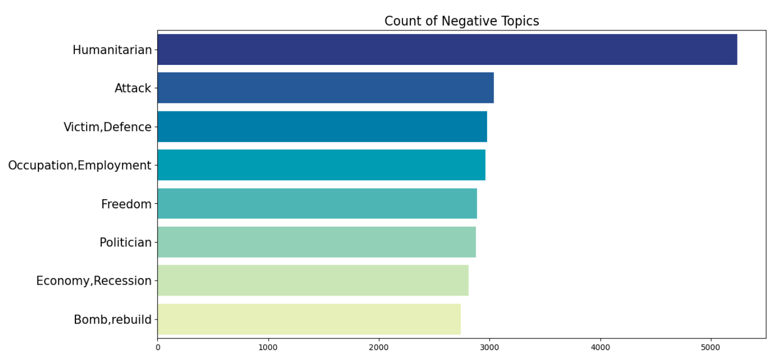
In this part, a topic modeling algorithm (LDA) is used to generate topics based on word frequency from tweet documents in part 2. Eight topics and their corresponding 25 most frequently occurring words for negative sentiments and positive sentiments are generated respectively.

3.1 Topics drive Negative sentiment

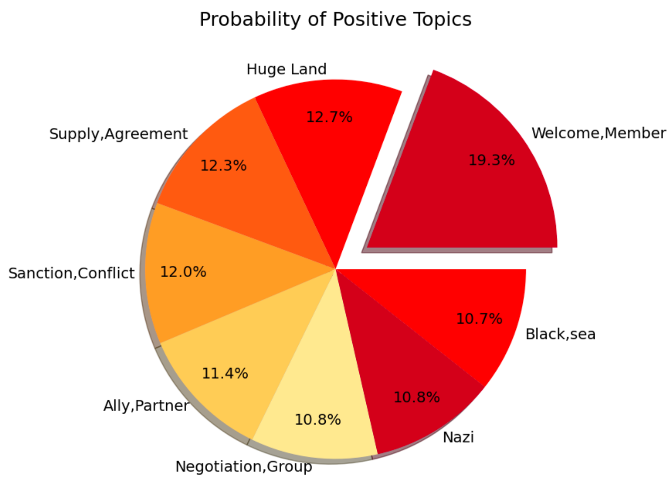


By looking at these words of negative sentiment in different topics in the pie chart (Figure 3-1), we recognize some representative word such as 'humanitarian', 'politics', 'murder', 'defending', 'freedom', 'shot', 'bombing'. From the pie chart, we find 20.5% of tweets care about humanitarianism which is the biggest group. 11.9% of tweets care about attack situations. 11.7% of tweets focus on victim and defense. So, we can see most people online pay attention to humanitarianism, attack, and victims.

People support peace and against war since Russia violates Ukraine's Sovereignty and humanitarianism. The war has brought harm and suffering to people, such as loss of freedom, destruction of buildings, and caused political chaos in Ukraine. In addition, there is an economic recession in Ukraine due to the double interference of war and pandemic. On the other hand, Russia and Ukraine are major producers of commodities, and disruptions in commodity production have led to sharp increases in global prices, especially oil and gas prices. Food costs have risen sharply, with wheat prices rising to record levels. Ukraine and Russia account for 30 percent of global wheat exports. Longer term, the war could fundamentally alter the global economic and geopolitical order if energy trade changes, supply chains are reshaped, payment networks fragmented, and countries reconsider their reserve currencies. Rising geopolitical tensions further increase the risk of economic fragmentation, especially over trade and technology. Thus, the economic problems in Ukraine will not only hurt Ukraine, but also may cause bigger problems involving the whole world.

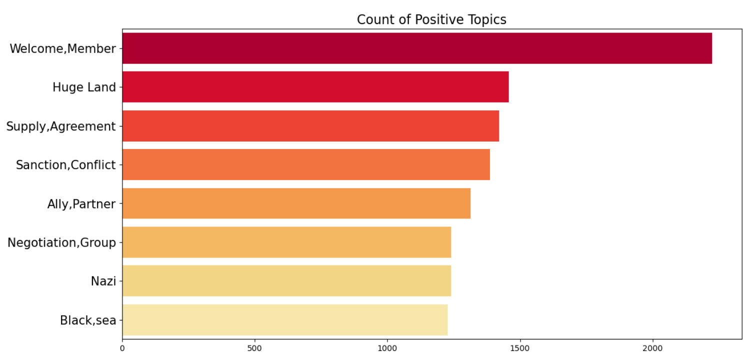


From the pie chart with 'Occupation and Employment', 'Bomb, rebuild', we can know that so many people lost their jobs because of war. Buildings were destroyed by Russian troops. People displaced and impoverished. The International Organization for Migration showed news to us that there are 60% of the displaced Ukraine people losing their jobs. The unemployment rate has risen to 34%, which doesn't capture the whole information because so many people in Ukraine are not recorded in the database.



3.2 Topics drive Positive sentiment

By looking at these words of positive sentiment in different topics, we recognize some representative words such as ‘member’, ‘welcome’, ‘huge’, ‘land’ in the topics. From Pie chart (Figure 3-3), we can see that approximately 20% of tweets with positive sentiments are related to the topic of ‘welcome and member’ and 12.7% of tweets are related to the topic of ‘huge land’ . Historically, Ukraine and Russia have the same root and origin, and the two countries belong to the East Slavic nation. Both Russia and Ukraine claim their heritage from Kievan Rus. Between this period and the collapse of the Soviet Union, both Russia and Ukraine are closely boned from one race and speak the same language. After the collapse of the Soviet Union, Ukraine became independent from the Soviet Union as a sovereign state. So many older Russians still see themselves in the Soviet era and consider Ukrainian territory as part of their own.



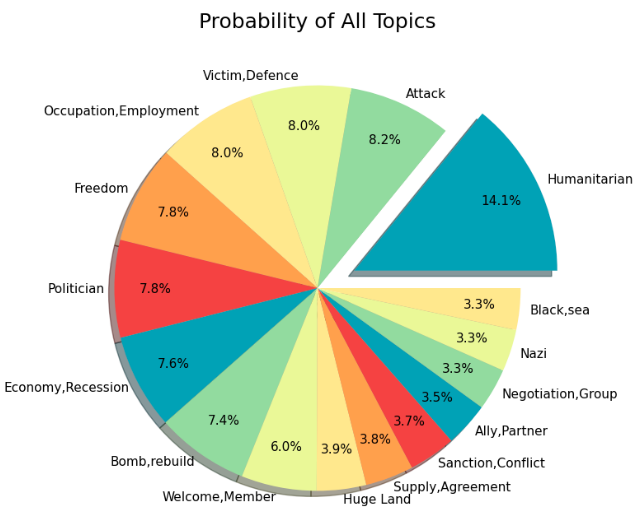
Some other obvious words such as ‘conflict’, ‘sanction’, ‘ally’, ‘partner’, ‘agreement’ can be related to some news reports. From the pie chart (figure 3-5), we can see that approximately 12% of tweets with positive sentiments are related to the topic of ‘sanction and conflict’ and 11.4% of tweets are related to the topic of ‘ally, partner’. From *Russians rightly unsettled by NATO's eastward creep* by Ehrhart, NATO has long maintained an "open-door policy" in the admission of members. Russia may be sensitive to NATO's eastward expansion plan, and as Ukraine applied for NATO membership, this may have made Russia perceive itself as threatened. From the pie chart, approximately 10.8 percent of tweets with positive sentiments are related to the topic of ‘Nazi’. The Russian government insisted on using the language of "special operations" and portrayed the campaign not as fighting the Ukrainian people, but as fighting the uncertain threat of “Nazism”, NATO. Some people whose tweets are classified as positive sentiment may tie Russia's invasion of Ukraine to the fight against Nazism, and try to protect their territory.

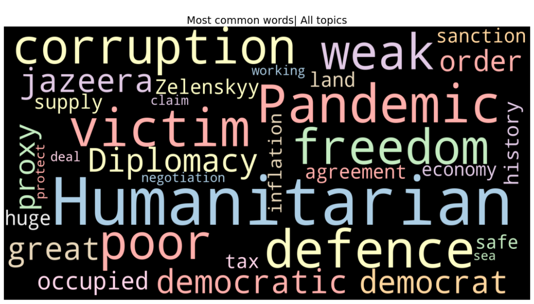
On the other hand, the pie chart also indicates 10.7 percent of tweets are related to ‘black and sea’. From the *Crimea and the Black Sea Fleet in Russian-Ukrainian Relations* by Victor Zaborsky, in the pre-war period, Ukraine and Russia argued endlessly over whose jurisdiction the Black Sea Fleet should be many years ago. This incident might sow the seeds of this war. Crimea is the home port of the Black Sea Fleet. Whoever controls Crimea can control the Black Sea Fleet and results in the control of the Black Sea Territory.

4 | Visualizations, Storytelling, Recommendations

4.1 Overall view of Topics

Based on the analysis of tweets of both positive and negative sentiment in part 3. As pie chart and word cloud (figure 4-1, figure 4-2) show more than 70 percent of tweets show negative sentiment towards the war. Among these topics of negative sentiment, most of the topics are about war harming human beings such as (Humanitarian, victim, freedom, attack etc). Before the war, the image of Ukraine as a Modern European Democracy. However, as the war progressed, people died and were displaced by the war, and the conflict raised serious humanitarian and democratic concerns.





4.2 Recommendation

Ukraine can publicize its domestic tragedy and displaced people's situation to the world, which can appeal to more people and more countries to focus on the Russia and Ukraine war. If Ukraine can get more attention and the world can realize what is happening in Ukraine, and Ukrainians really need help, which is helpful for them to receive more support.

To reduce the harm of conflicts to civilians, call on parties to the conflict to exercise maximum restraint, strictly abide by international humanitarian law, protect the safety of civilians and civilian facilities, and facilitate the evacuation of personnel and humanitarian aid operations. The parties should be concerned about strengthening communication on humanitarian issues, and conduct coordination and cooperation, such as opening humanitarian corridors and organizing the safe evacuation of personnel. We support Secretary-General Guterres and Deputy Secretary-General Griffith in contacting all parties on the humanitarian ceasefire issue and reaching a consensus on relevant arrangements.

The refugee issue must be properly dealt with. The countries surrounding Ukraine have provided safe shelter and humanitarian assistance to millions of refugees. The continuation of the conflict will give birth to a larger refugee group and bring some economic and social challenges to refugee-hosting countries. Ukraine can show its thanks to countries that offer help during war in the media. Besides, action is needed to crack down on human trafficking and other crimes targeting refugee women and children. UNHCR, UN Women, UNICEF, OHCHR and other institutions should also strengthen monitoring and support the efforts of relevant countries.

Ukraine needs to show the world that Ukraine is not the only one being hurt. There are still many sub consequences that will spill over to other countries. By *Commodity price increase due to the Russia-Ukraine war 2022* from Statista, the global wheat prices had increased by over 60% during the period from February 24 to June 1, 2022. The growth was explained by the Russia-Ukraine war, as Russia and Ukraine were among the leading wheat exporters. Furthermore, coal prices grew by around 69 percent. A significant increase was also recorded in prices of metals exported by Russia, such as nickel, palladium, and aluminum. Those statistics can make the public realize the wider impact of the war negatively.

Finally, increase the sense of urgency to promote diplomatic negotiations. The prospect of expanding and protracted conflicts is worrisome. Cease-fire and cessation of war as soon as possible is the fundamental way out to resolve the humanitarian crisis. We call on Russia and Ukraine to stick to the general direction of dialogue and negotiation, continue to narrow differences, and accumulate conditions for the realization of a ceasefire.

6 | Reference

- Gawade, O. (2022, March 16). Russia|Ukraine war — tweets NLP analysis. Medium.https://omkargawade.medium.com/russia-ukraine-war-tweets-nlp-analysis-bd10b352316c
- Russia/Ukraine conflict: UQ experts. UQ News. (2022, February 25). Retrieved December 3, 2022, from https://www.uq.edu.au/news/article/2022/02/russiaukraine-conflict-uq-experts
- Rott, N., Harbage, C., & Malofieieva, K. (2022, November 2). Millions of Ukrainians have escaped the war. many still can't find enough work. NPR. Retrieved December 4, 2022, from (..., 2019)https://www.npr.org/2022/11/01/1132167234/russia-ukraine-war-unemployment-displaced-economy
- Published by Statista Research Department, & 24, N. (2022, November 24). Commodity price growth due to Russia-ukraine war 2022. Statista. Retrieved December 4, 2022, from https://www.statista.com/statistics/1298241/commodity-price-growth-due-to-russia-ukraine-war/
- Los Angeles Times. (2022, November 21). Op-ed: Negotiations can't end the Russia-Ukraine War until one side has lost. Los Angeles Times. Retrieved December 4, 2022, from https://www.latimes.com/opinion/story/2022-11-21/ukraine-russia-war-negotiated-peace-cease-fire
- Published by Statista Research Department, & 24, N. (2022, November 24). Commodity price growth due to Russia-ukraine war 2022. Statista. Retrieved December 4, 2022, from https://www.statista.com/statistics/1298241/commodity-price-growth-due-to-russia-ukraine-war/
- Los Angeles Times. (2022, November 21). Op-ed: Negotiations can't end the Russia-Ukraine War until one side has lost. Los Angeles Times. Retrieved December 4, 2022, from https://www.latimes.com/opinion/story/2022-11-21/ukraine-russia-war-negotiated-peace-cease-fire
- Ehrhart, WD. “Guest Opinion: Russians rightly unsettled by NATO's eastward creep.” Bucks County Courier Times, 20 February 2022, https://www.buckscountycouriertimes.com/story/opinion/2022/02/20/russians-rightly-unsettled-natos-eastward-creep/6832991001/. Accessed 4 December 2022.
- Lynn, S. M., Nye, J. S., & Brooks, H. (n.d.). (1995, September). Crimea and the Black Sea Fleet in Russian- Ukrainian Relations. Belfer Center. Retrieved December 4, 2022, from https://www.belfercenter.org/publication/crimea-and-black-sea-fleet-russian-ukrainian-relations